

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pesatnya perkembangan dunia informasi dan teknologi, memungkinkan tersedianya informasi yang melimpah sehingga menyebabkan kebutuhan informasi akan suatu hal juga meningkat. Hal ini terjadi disemua Negara, Indonesia salah satunya, kebutuhan informasi khususnya informasi berbahasa Indonesia semakin lama semakin meningkat sehingga menyebabkan riset – riset terhadap mesin pencari khususnya mesin pencari untuk informasi berbahasa Indonesia.

Search engine atau mesin pencari yang merupakan salah satu penerapan dari *Information Retrieval (IR)* dapat memaksimalkan hasil dari penghimpunan *web crawler* sehingga memungkinkan pengguna untuk mencari informasi dari data yang telah terhimpun hanya dengan memasukkan sebuah *query* yang diinginkan, maka data yang relevan terhadap *query* akan ditampilkan berdasarkan tingkat kemiripan yang paling besar.

Pencarian informasi dengan *query* pada sistem *IR* memang sangat berguna dalam memberikan daftar berita dengan hasil yang maksimal dan relevan terhadap *query* yang dimasukkan oleh pengguna, namun pencarian dengan *query* biasa ini tidaklah efektif dikarenakan hasil berita yang muncul pada peringkat awal belum tentu relevan terhadap keinginan pengguna. Jadi, dibutuhkan adanya suatu pengelompokan terhadap data-data yang memiliki isi atau topik yang mirip untuk mengatasi kendala tersebut.[1] Pengelompokan data ini bertujuan agar lingkup pencarian dapat dipersempit berdasarkan kategori – kategori yang telah terbentuk, sehingga hanya dokumen yang sejenis yang akan ditampilkan.

Telah banyak metode pengelompokan data yang dikembangkan [2], *Pretopology* adalah salah satunya. Penerapan metode *Pretopology* dengan menggunakan model ruang vektor (*Vector Space Model*) untuk merepresentasikan kemiripan antar berita, dapat mengelompokkan data sesuai dengan kategori – kategori yang ada secara maksimal, sehingga kualitas pencarian akan menjadi lebih relevan sesuai dengan keinginan pengguna, selain itu dengan penggunaan metode *pretopology* yang menitikberatkan terhadap penggunaan representasi dokumen terhadap

setiap kelas yang ada, menyebabkan proses pengelompokan data menjadi lebih cepat.

Tugas akhir ini akan membangun sebuah sistem klasifikasi teks berbasis *web search engine* terhadap berita yang telah dihimpun oleh mesin *crawler* dengan menerapkan metode *Pretopology* dalam pengelompokan data. Diharapkan dengan adanya penelitian ini, data yang telah terhimpun oleh mesin *crawler* dapat digunakan secara maksimal dan pencarian informasi dapat menjadi lebih dipersempit berdasarkan kategori – kategori yang ada, sehingga sistem dapat memberikan hasil berita yang lebih relevan terhadap keinginan pengguna.

1.2 Perumusan Masalah

Beberapa hal yang menjadi permasalahan dalam tugas akhir ini adalah sebagai berikut:

1. Bagaimana cara meningkatkan kualitas hasil pencarian pada mesin pencari?
2. Bagaimana klasifikasi dapat meningkatkan kualitas hasil pencarian?

1.3 Tujuan dan Manfaat

Tujuan dari penelitian ini adalah membangun sebuah sistem klasifikasi berita dengan menggunakan metode *Pretopology* dalam pengklasifikasian data. Sehingga dengan adanya pengklasifikasian terhadap data yang ada, pengguna dapat melakukan pencarian data berdasarkan kategori – kategori yang telah terbentuk dalam klasifikasi sehingga kualitas hasil pencarian pada mesin pencari dapat ditingkatkan.

1.4 Batasan Masalah

Demi tercapainya tujuan penelitian dengan baik, maka perlu adanya batasan masalah penelitian, yaitu:

1. Tidak menangani pembuatan mesin *crawler*. Mesin *crawler* yang digunakan adalah *vietspider 3.17*.
2. Data hanya terbatas pada hasil penghimpunan mesin *crawler*.
3. Website yang di *crawler* adalah web portal berita Kompas.com

dengan 8 sub URL, yaitu:

- <http://olahraga.kompas.com>
- <http://bola.kompas.com>
- <http://tekno.kompas.com>
- <http://health.kompas.com>
- <http://edukasi.kompas.com>
- <http://travel.kompas.com>
- <http://entertainment.kompas.com>
- <http://properti.kompas.com>

Total berita yang diambil dari 8 sub URL diatas adalah 9172 berita.

4. Klasifikasi dokumen menggunakan metode *Pretopology*.
5. Kategori yang digunakan untuk klasifikasi adalah kategori olahraga, teknologi, kesehatan, pendidikan, travel, entertainment, dan properti.

1.5 Metodologi

Pembuatan Tugas Akhir ini dilakukan dengan menggunakan metodologi sebagai berikut :

1. Studi Pustaka

Mengumpulkan data dengan cara mencari dan membaca buku sebagai bahan literatur untuk membantu pemecahan masalah mengenai permasalahan pengkategorian dokumen secara umum, *Vector Space Model* yang digunakan pada Information Retrieval, serta penerapan metode *Pretopology* dalam pengkategorian dokumen.

2. Perancangan Sistem

Tahap ini meliputi analisis dan desain sistem yang akan dikembangkan dengan mengacu pada hasil pemahaman sistem dan studi literatur serta data yang telah diperoleh sebelumnya.

Pada tahapan ini, desain model data, dan desain proses didefinisikan sehingga dapat memenuhi kebutuhan sistem.

3. Implementasi Sistem

Pada tahap ini dilakukan pembuatan program berdasarkan deskripsi dan alur sistem yang telah dibuat.

4. Evaluasi Sistem

Melakukan pengujian terhadap aplikasi yang baru agar diketahui kesalahan dan dapat diperbaiki kembali untuk mendapatkan hasil yang terbaik.

5. Dokumentasi

Pada tahap ini dilakukan penulisan mengenai apa yang diperoleh dan dikerjakan mulai dari kegiatan studi pustaka hingga menghasilkan aplikasi yang baru.

1.6 Sistematika Penulisan

Adapun sistematika pembahasan tugas akhir ini adalah sebagai berikut :

BAB I : PENDAHULUAN

Pada bagian ini terdiri dari : latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, metode penelitian dan sistematika penulisan.

BAB II : TINJAUAN PUSAKA

Pada bagian ini dijelaskan mengenai hal-hal yang melandasi atau yang akan digunakan sebagai pedoman untuk menyelesaikan permasalahan yang terdapat dalam tugas akhir ini.

BAB III : PERANCANGAN SISTEM

Pada bagian ini akan dijelaskan pemodelan dan jaringan yang akan digunakan dalam melakukan penelitian tugas akhir.

BAB IV : SIMULASI DAN ANALISIS

Pada bab ini menjelaskan tentang implementasi program yang telah dikerjakan. Meliputi penjelasan tentang arsitektur sistem, desain sistem, desain interface, dan menguraikan tentang hasil uji coba serta analisa terhadap sistem.

BAB V : PENUTUP

Bab yang berisi kesimpulan dan saran yang dihasilkan dari sistem yang telah diujicobakan.

[HALAMAN INI SENGAJA DIKOSONGKAN]